DOCUMENT RESUME

ED 333 018 TM 016 468

AUTHOR Chou, Tungshan; Huberty, Carl J.
TITLE A Freshman Admissions Prediction Equation: An Evaluation and Recommendation.
PUB DATE 90
NOTE 43p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations; *College Freshmen; Decision Making; *Equations (Mathematics); Grade Point Average; Grouping (Instructional Purposes); Higher Education; *Mathematical Models; Predictive Measurement; Predictive Validity; *Predictor Variables; Racial Differences; Sex Differences
IDENTIFIERS *Freshman Admissions Prediction Equation; Scholastic Aptitude Test; University of Georgia

ABSTRACT

The effectiveness of the freshman admissions prediction equation (FAPE) at the University of Georgia (Athens) was studied, using data for 3,378 freshman (1,490 males and 1,888 females) in 1987-88. For the 1987-88 data, a previous (1982) prediction equation functioned poorly in terms of predictive accuracy. New prediction models were constructed using multiple regression techniques and analyzed in terms of: mean residual (average prediction error), mean absolute residual (average value of absolute prediction errors), standard deviation of the residuals (standard deviation of prediction errors), and misclassification rates (proportion of students who did and did not perform acceptably given that they were predicted not to and to do well, respectively). The predictors used include: verbal and mathematics scores on the Scholastic Aptitude Test, high school grade point average, gender, race, and high school grouping. The 1986-87 data for 3,177 freshmen (1,353 males and 1,824 females) were used to validate the effectiveness of two prediction models. Inclusion of gender, race, and high school grouping did not improve the predictive accuracy in terms of mean absolute residual, residual standard deviation, and misclassification rates; however, some improvement in reducing the mean residual was observed. Since some previous studies show that the effectiveness of predictors varies with institutional characteristics, the construction/implementation of the prediction model should consider gender, race, and high school grouping factors. Two ways of increasing a prediction equation's utility, and recommendations for FAPE users are proposed. Eleven tables, one graph, and an 18-item list of references are included. (RLC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TUNGSHAN CHOU

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED333018

A Freshman Admissions Prediction Equation: An Evaluation and Recommendation

Tungshan Chou Carl J Huberty

The University of Georgia

325 Aderhold Hall
Department of Educational Psychology, Research, & Measurement
The University of Georgia
Athens, GA 30605

TM 016 482

BEST COPY AVAILABLE

Abstract

This study investigated the effectiveness of the freshman admissions prediction equation at The University of Georgia. A previous prediction equation used was found to function poorly in terms of predictive accuracy, and new prediction models were constructed using multiple regression techniques on the 1987-88 freshman data. Four criteria were considered in the evaluation of the new models: (1) mean residual, (2) mean absolute residual, (3) standard deviation of the residuals, and (4) misclassification rates. The predictors under consideration included Scholastic Aptitude Test Verbal and Mathematics scores, high school grade point average, gender, race, and high school grouping. The 1986-87 freshman data were used to validate the effectiveness of two prediction models. It was found that the inclusion of gender, race, and high school grouping did not improve the predictive accuracy in terms of mean absolute residual, residual standard deviation, and misclassification rates; some improvement in reducing the mean residual was observed, however.

Even though the inclusion of gender, race, and high school grouping factors did not improve the predictive accuracy at The University of Georgia, there is some evidence from previous studies that the effectiveness of the predictors varies with institutional characteristics. Therefore, the construction and implementation of the prediction model should take into account these characteristics. Finally, two ways of increasing the utility of an established prediction equation were suggested, and several recommendations for the users of the freshman admissions prediction equation are advanced.

A Freshman Admissions Prediction Equation: An Evaluation and Recommendation[1]

The effectiveness of using a prediction equation based on some prior college predictors in making freshman admission decisions has been questioned by some researchers in the past. Predicting student academic achievement independently of what the college does to assist or to constrain achievement is judged by some to be highly problematic. Being confronted with a large number of applicants, however, most higher education institutions choose to use some sort of prediction equation to help make admission decisions. The use of a prediction equation may be justified if its use is better than any other feasible method of selecting prospective students. However, evaluation of such an equation should not solely focus on how accurately it predicts an individual's academic performance. Rather, the evaluation should also focus on its utility in making admission decisions.

A large body of research has been done in search for good predictors of early academic failures (Breland, 1979; Ott, 1988; Terenzini, 1982). The effectiveness of the predictors has traditionally been assessed by the correlation between a composite of the predictors and a college grade point average.

---

1. The significant contributions of professors James Fortson, Jeremy Kilpatrick, and John Neter (all three at The University of Georgia) toward early suggestions pertaining to prediction models and analyses are acknowledged and greatly appreciated.

Three pre-college achievement measures relevant to early academic failures (in most instances, academic failure means a grade point average of less than 2.0 on a 0-4 scale) identified in these studies were the Scholastic Aptitude Test Verbal (SATV) and Mathematics (SATM) scores, and high school grade point average (HSGPA). Baird (1983) reported the multiple correlation of these three measures with the first year college grade to be .56 in general, without adjusting for restriction of range. Weitzman (1982) applied statistical adjustment for the restriction of range effect, and reported the multiple correlation of college grade point average with a composite SAT score and HSGPA to be .76. However, Baird (1983) found that the predictability of college performance varies with different institutional characteristics, such as the percentage of minority students, commuters versus non-commuters, diversity of programs, etc. Ott (1988) investigated the effect of two demographic variables, race and gender, on the academic failure after controlling for the pre-college achievement variables, and made two conclusions: (1) race is significantly related to early academic failure; (2) there is no significant difference in the predicted probability of academic failure between men and women.

Some studies of selection process investigated the "relative importance" of predictors in the decisions of selecting employees or college admissions (Arvey & Faley, 1988; Gatewood & Feild, 1987; Hunter & Hunter, 1984; Russell,

Persing, Dunn, & Rankin, 1990). Researchers have generally concluded that cognitive assessments have higher validities in the selection process than most other measures (Hunter & Hunter, 1984). These studies also indicated that the predictors differ in their usefulness, and therefore, should not be given equal weight in making selection decisions (Arvey & Faley, 1988; Hunter & Hunter, 1984; Russell et al., 1990). However, in contrast to Ott (1988), Russell et al. (1990) reported gender of minor importance, and race was non-significant in the admission process. For most of the studies mentioned, the use of regression techniques is a popular way of making selection predictions. Path analysis (Pedhazur, 1982) and structural equation modeling (Joreskog & Sorbom, 1986) are two extensions of regression techniques which are often employed in the comfirmation of the proposed models, or identification of relative predictor "importance" in the selection process. The presently discussed prediction equation for college admissions is also an application of the regression technique which attempts to identify likely future academic failures by means of estimating predicted college grade point average.

Most studies dealing with the evaluation of college grade prediction have focused on the strength of relationship (as an index of goodness of prediction) between the pre-college achievement measures and the actual obtained college grade point average (Baird, 1983; Educational Testing Service, 1980; Ford & Campos, 977; Weitzman, 1982). The issue of how effective

these pre-college measures can be to help make admission decisions needs to be further addressed in terms of mean square error and the utility of these measures for making admission decisions. The mean square error gives an indication of the likely error range in prediction. The utility of predictive measures should be assessed in terms of their ability to identify "successful" and "failing" students. Because there is an arbitrary cut-off point for "successful" or "failing," the evaluation of the prediction model is dependent upon this arbitrary decision. Given an established prediction equation, a pertinent question may be: What predictive score should the administrators determine as the cut-off score for making admission decisions so that the utility of this prediction equation is maximized? Or, given a cut-off score, what is the likely error rate for this prediction equation?

The current study investigated the effectiveness of the freshman admissions prediction equation at The University of Georgia, a major teaching and research university in the Southeast with a wide academic program diversity. An initial objective was to determine the acceptability of using a 1982 prediction equation on 1987-1988 incoming freshmen. It was found that this prediction equation functioned poorly with the 1987-88 freshman class. Efforts were undertaken to construct new models which could result in better predictive accuracy. We investigated the effectiveness of the prediction model with SATV, SATM, and HSGPA as the predictors. We also constructed

prediction models which included race, gender, and high school grouping as predictors in addition to the two SAT scores and HSGPA. The effects of including gender, race, and high school grouping in the prediction model on the predictive accuracy were assessed according to four criteria for the evaluation of prediction models. This paper summarizes the findings from our study and makes recommendations for implementing a prediction equation so that the utility of such an equation may be increased. All analyses were done using the Statistical Analysis System (SAS, 1985).

Previous Model

A prediction model used until quite recently was developed on the 1982-83 freshman class. The criterion measure was fall quarter college grade point average (FCGPA) while the predictors are SATV, SATM, and adjusted high school grade point average (adj. HSGPA). The latter measure involved a numerical value (.00, .20, .30, .40, .50, or .60) -- label it HSGPA adj. factor -- added to the HSGPA. The amount added for a given student was based upon prior prediction of FCGPA, if there were at least 10 students from that particular student's high school. To the extent that the prediction differed from the actual mean performance value, an adjustment was made.

The 1982-83 equation was applied to a total group of 3818 1987-88 incoming freshmen, and to 17 subgroups: (1) male, (2)

female, (3) admission status 1 (regular), (4) admission status 5 (committee admit), (5) admission status 6 (developmental studies), (6) white, (7) non-white, (8) black, (9) non-black, (10) HSGPA adj. factor = .00, (11) HSGPA adj. factor = .20, (12) HSGPA adj. factor = .30, (13) HSGPA adj. factor = .40, (14) HSGPA adj. factor = .50 or .60, (15) $\hat{Z} \leq 2.30$, (16) $2.30 < \hat{Z} < 2.80$, and (17) $2.80 \leq \hat{Z}$ ($\hat{Z}$ denotes the predicted FCGPA).

The results of applying the 1982-83 equation to the 1987-88 data revealed large biases for some subgroups; e.g., for the 313 Developmental Studies students, the mean residual was as large as -1.65, while the other subgroup mean residuals were in the -.55 to -.20 range. In addition, the variability of prediction errors (as measured by the standard deviation of the residuals) was judged to be large, in the range of .70 to .80. Based on the results of these analyses, two conclusions were made: (1) a more appropriate (in the sense of mean predictive accuracy) criterion measure would be the cumulative grade point average over fall, winter, and spring quarters (CCGPA) for students earning a total of at least 25 quarter hours; and (2) Developmental Studies students data should not be used in constructing a final prediction rule.

Constructing New Models

In a search for a better prediction model, we constructed a number of preliminary models that included different

combinations of predictors. A total of 11 models and 16 predictors were considered. In addition to the predictors used in the 1982-83 model, we included gender, race, the unadjusted HSGPA, adj. HSGPA, and HSGPA adjustment factor. The latter predictor was considered in two ways, as a quantitative variable and as a grouping variable. To include adj. HSGPA factor as a grouping variable in the regression analysis, we used a dummy coding scheme to code four dummy variables. These 11 models were then applied to 3378 1987-88 incoming freshmen, excluding 313 Developmental Studies students and those with less than 25 cumulative quarter hours, and to the 16 subgroups as defined previously. We also included a squared transformation of the criterion variable as well as some cross-product terms in some of these 11 models. The variable definitions and the models are listed in Table 1.

Insert Table 1 about here

Models A and B were first constructed and compared to assess the possibility of a non-linear relationship between the predictors and the college performance. Model C used three cognitive measures as predictors: SATM, SATV, and unadjusted HSGPA. These measures were judged by some researchers as the most valid measures in the selection process (Hunter & Hunter, 1984; Russell et al., 1990). The results of analyses based on

Model C showed moderately large biases (the residual means exceeded .10) existing in some HSGPA adjustment factor subgroups and for blacks. The gender bias was also present to a minor extent (the residual means are -.046 and 0.036 for males and females, respectively). Models D through G were considered to investigate the effect of adding HSGPA adj. factor as a grouping variable and/or some interaction terms in the prediction equation. Models H and I were able to remove the gender and race biases; there still remained some bias for the HSGPA adjustment factor subgroups, however. Finally, we grouped 85 high schools into five groups based on their average residual errors from negative to positive. These 85 high schools, each of which sent at least 10 1987-88 freshmen who completed the academic year with at least 25 quarter hours credit, accounted for about 2100 students. The remaining 1400 or so students came from some 670 other high schools, and they constituted the sixth group. The six groups were determined so that the numbers of students in the first five groups were approximately equal. These six groups lead to the consideration of five indicator variables (X12 through X16) as additional predictors. Model J consisted of SATV, SATM, (unadjusted) HSGPA, gender, race, and five indicator variables for defining the six high school group memberships.

The results of applying Model J to the total sample and the subgroups indicated that the prediction bias is virtually nonexistent for most of the subgroups. The predicted CCGPA and

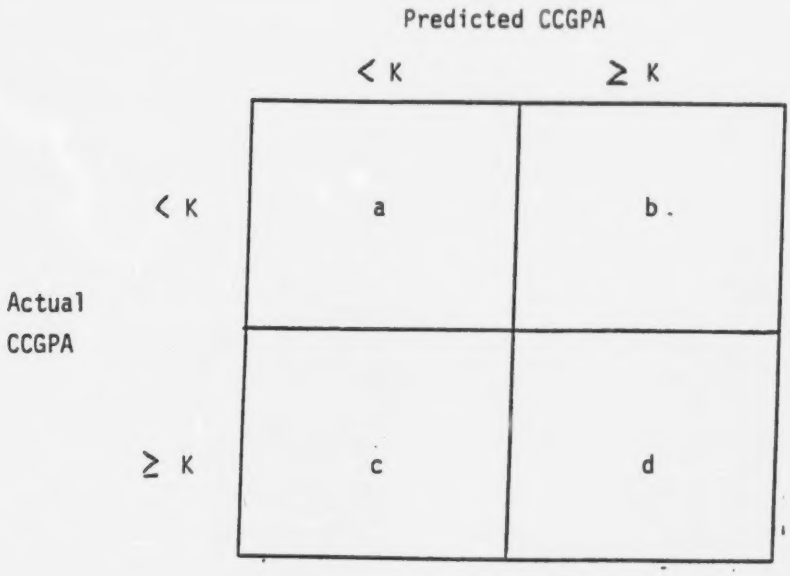
the actual CCGPA were, on the average, the same. The absolute residual means were roughly between .40 to .45. The standard deviation of the residuals was, across the subgroups, approximately .53. It turned out that the five cross-product variables in Model K did not contribute to increased predictive accuracy over that yielded by Model J. Therefore, our interest centered on Model J. For illustrative and comparative purposes, we report the results from Model C and Model J with respect to four criteria of evaluation.

Four Criteria for Model Assessment

In assessing the goodness of the 11 models, we considered four criteria: (1) mean residual (average prediction error), (2) mean absolute residual (average value of the the absolute prediction errors), (3) root mean square error (the standard deviation of the prediction errors), and (4) misclassification rates (proportion of students who perform acceptably given that they are predicted not to do well, and proportion of students who do not perform acceptably given that they are predicted to do well). These measures were examined in a comparative sense for the total group and the subgroups. For the comparison of Model C and Model J, it was decided to focus on the total group and on 13 subgroups: male, female, black, non-black, six HSGPA adjustment factor subgroups, $\hat{Y} < 2.30$, $2.30 \leq \hat{Y} \leq 2.80$, and $2.80 < \hat{Y}$ ($\hat{Y}$ denotes the predicted CCGPA). The lower the value

of each of these four criteria, the better is the model.

The fourth assessment of a prediction model involves a direct, practical assessment of the predictive accuracy of the model. It involves determining a CCGPA value K, for example, K=2.00, as a minimum level of "satisfactory" performance. The selected K value is used as a cut-off point for predicting success or failure. A 2*2 classification table which shows the hits and misses may be constructed. The ratio of misses to total N can be used as an index of overall prediction accuracy. In addition to the overall prediction accuracy index, we also considered two types of error rate indices: false-positive and false-negative. The generic table is represented as below:



The cells with "b" and "c" indicate misses whereas the cells with "a" and "d" indicate hits. Cell b represents false-positive errors, when students are predicted to do well but turn out to be low achievers; and cell c represents false-negative errors, when students are predicted not to do well but actually achieve at an acceptable level. The false-positive and false-negative error rates reported in this study are conditional on the numbers of people predicted to be in the "success" and "failure" groups, respectively; and they do not sum to the total error rate. The inherent problem confronting researchers in evaluating the predictive accuracy of a prediction model is the unbalanced nature of the group dichotomy: only a small proportion of people would be dichotomized into a "failure group." It is understandable that applicants with low scores on the predictive measures are unlikely to be admitted just for the purpose of prediction model evaluation. The "failure" group may in fact be underrepresented since many students have not been admitted because of low predicted CCGPA upon applcation. Some might argue that the false-negative error rate is the more serious one. For a better assessment of false-negative errors, one may need to redefine the cut-off point for the "success" and "failure" group for the purpose of model evaluation. Obviously, the selection of K is crucial in determining these misclassification error rate indices.

Tables of the type illustrated here can be constructed for

any K value where K is the arbitrary cut-off CCGPA which is considered as the minimum level of satisfactory performance. Such classification tables may be useful in assessing the predictive accuracy for a particular prediction model, or for comparing accuracy across models. Of course, values of K are still subjective administrative decisions. However, the usefulness of the prediction equation can now be evaluated in terms of these administrative decisions. For initial discussion purposes, we chose a K value of 2.00. Later we will also see how the error rates change as we change the value of K.

Model C Using 1987-88 Data

The descriptive information of the three predictive measures and CCGPA for the 1987-1988 freshmen is summarized in Table 2. The correlation of CCGPA with HSGPA is .52, which is

---

Insert Table 2 about here

---

the same as the predictive validity of HSGPA reported by Slack and Porter (1980). The predictive validities corrected for the restriction of range effect for HSGPA and the composite SAT score were reported by Weitzman (1982) as .60 and .64, respectively. Weitzman also showed that the combination of HSGPA and the composite SAT score lead to greater predictive

validity (.76, adjusted for restriction of range) than either one used alone. It is believed that the inclusion of both SATV and SATM scores in the prediction equation would further increase the predictive validity.

Regression diagnostics showed no indications of multicollinearity (the variance inflation factors range from 1.09 to 1.26). The number of outliers as identified by the studentized residuals greater than 3 in absolute value is about 30, roughly what one would expect by chance. These outliers were judged not to be influential. A summary of the internal residual analyses for this model for the total group and the 13 subgroups is given in Table 3. An internal analysis is one for

---

Insert Table 3 about here

---

which the results are based on applying the prediction rule to the same data that were used in deriving the rule. The residual means showed moderately large biases for four subgroups: Black and the last three HSGPA adjustment groups. For the three HSGPA adjustment groups, Model C underestimated applicants' actual college academic performance, whereas for the blacks, actual college performance was being overestimated. The absolute residual means and the residual standard deviations are indications of the prediction error variability. The larger the values of absolute residual means or the residual standard

deviations, the less precise the prediction model is. The error variabilities are similar for all subgroups.

The misclassification error rates are summarized in Table 4. From the table we see, for example, about 19.3% of the 1987-

---

Insert Table 4 about here

---

88 males who were predicted to have a CCGPA greater than 2.00, had actual CCGPA less than 2.00. On the other hand, 47.0% the males who were predicted to have a CCGPA less than 2.00, had actual CCGPA greater than 2.00. Of the 1490 males, a total of 320 (21.48%) "misses" resulted from using the prediction equation. There should not be a total surprise on the 100% false-negative error rate for the HSGPA adjustment factor .00 group, since there was only one person predicted to fail and this individual turned out to be successful. For all of the subgroups and total group, the false-negative error rates are judged to be unacceptably high. Because the total number of people predicted to be in the "failure" group is proportionally small (10% to 20%) compared to the number of people predicted to succeed, the large false-negative error rates did not appreciably affect the overall classification accuracy of the model. However, these high false-negative error rates may be a great concern to some administrators. Misclassification rates will be discussed further in subsequent sections.

Model J Using 1987-88 Data

Model J takes consideration of gender, race, and high school grouping variables. Since the largest minority population at UGA is blacks (n=162, less than 5% of the student body), and other races are very few, the race variable was used to classify blacks and nonblacks. Two students did not indicate racial affiliation, and thus were dropped from the analyses; leaving a total number of 3376 cases for Model J construction. The inclusion of these additional variables is only marginally helpful here -- the root error mean square decreased from .55 to .53, and the adjusted R^2 increased from .33 to .37. However, the grouping variables may be more helpful when the groupings are based on cumulative data consisting of several years. This will be discussed in the following section. No indications of multicollinearity and influential observations were found with this model.

A summary of the internal residual analyses for the total group and the 13 subgroups is given in Table 5. For most of the 13 subgroups, the prediction bias is virtually nonexistant;

Insert Table 5 about here

---

i.e., the predicted CCGPA and actual CCGPA were, on the

average, the same. The absolute residual means were roughly between .36 and .48 with nine of the 13 subgroups having values in the .40 - .44 range. The standard deviation of the residuals is, across the subgroups, approximately .53 (which is roughly the same as the root error mean square); i.e., one can be about "67% sure" that a student's actual CCGPA will be within .53 of his or her predicted CCGPA score. Comparing Table 5 to Table 3, the inclusion of gender, race, and high school grouping eliminated the biases in terms of mean residuals. However, there is no advantage of Model J over Model C in terms of absolute residual mean and residual standard deviation. The results of misclassification error rates are summarized in Table 6. Comparing Table 6 to Table 4, it appears that Model J does not show significant improvement in terms of misclassification rates either.

Insert Table 6 about here

Validating Prediction Models with 1986-87 Data

Of course, data that were used to construct the prediction model should not be used again to evaluate its effectiveness since such an internal analysis would most likely reveal the best results with such a specific sample. An external analysis

which applies the rules of the constructed models to a different freshman class is preferred. We employed the 1986-87 freshman data for validation purposes. The descriptive information for the 1986-87 data is reported in Table 7.

Insert Table 7 about here

Model C Validation

The weights of the established rule were the coefficients of the regression equation determined by using 1987-88 freshman data. Applying this rule to the 1986-87 freshman data and the predicted CCGPA were obtained. The residual analyses for the total sample and the subgroups are shown in Table 8. Comparing Table 8 to Table 3, it appears that Model C rule holds similar

---

Insert Table 8 about here

---

results for 1986-87 data as for the 1987-88 data in terms of the mean residuals. With respect to absolute residual means and residual standard deviations, Model C shows better results for 1986-87 data than for the 1987-88 data. The misclassification error rates are reported in Table 9. Comparing Table 9 to Table

Insert Table 9 about here

4, the overall error rates, for the total group as well as for all subgroups, are less than those reported for the 1987-88 data. The false-positive error rates for most subgroups for the 1986-87 data are less than those reported for 1987-88 data. The false-negative error rates, however, are generally slightly higher than those for 1987-88 data.

Model J Validation

To accomplish validation for Model J, the high schools representing the 1986-87 students were placed into the same groups as defined by the 1987-88 data. The external residual analyses are reported in Table 10. Comparing Table 10 to Table 5, Model J also seems to hold similar results across years.

Insert Table 10 about here

Nevertheless, Model J still yielded moderately large bias for the black subgroup. A comparison of Tables 10 and 8 indicates similar absolute residual means and residual standard deviations using either Model C or Model J. Model J shows slight advantage over Model C in terms of residual means for

some HSGPA adjustment factor subgroups. However, some biases still resulted for some HSGPA adjustment factor subgroups. This may not be too surprising since high school groupings based on the results for a single year are subject to erratic changes. If the grouping of high schools is done based on cumulative years data, the biases among different high schools will be minimized, and therefore there will remain no need for the HSGPA adjustment factor in computing the prospective students predicted CCGPA. The misclassification rates were given in Table 11, and are similar to those reported for 1987-88 data.

Insert Table 11 about here

Summary

Both Model C and Model J hold similar results over years in terms of all four criteria. The inclusion of gender, race, and high school grouping slightly reduced the mean residuals for male, female, black, and some HSGPA adjustment factor subgroups. However, it did not improve the predictive accuracy in terms of mean absolute residual and standard deviation of the residuals. These results are consistent with studies which reported high predictive validities for the cognitive measures, but not for gender and race (Baird, 1983; Hunter and Hunter, 1984). With respect to the utility of the prediction equation for making admission decisions, using CCGPA of 2.00 as the cut-

off point for defining "success" or "failure," the inclusion of gender, race, and high school grouping does not help reduce the misclassification error rates. Nevertheless, it should be pointed out that the black population may be underrepresented in our study since only approximately 5% of the freshmen in both years were blacks.

Given the predictive measures we have, the results of evaluation for the two prediction models may seem rather disappointing. Many of the total error rates are approximately .20; that is, about one out of five students was misclassified in the total sample and in many subgroups. Additional predictive measures will be needed if one wishes to improve the effectiveness of the prediction model. Although Model J did not show significant advantage over Model C, there is some evidence from previous studies, however, that the predictive validities for the cognitive measures used in this study may vary with institutions of different characteristics (Baird, 1983; Willingham, 1990). For the administrative purposes, we discuss the utility of the prediction equation using Model J for the remainder of this paper.

Implementing Model J

The biggest problem with implementing Model J is the grouping of the high schools; that is, how to assign values of X_{12} through X_{16}. The size of group six, as determined from the

1987-88 data alone, is some 3.5 times the sizes of the other five groups. To achieve greater group size equality as well as more reliable high school mean prediction residuals, it is suggested that a "moving average" approach be used to arrive at the mean prediction residuals. This approach is described as follows.

The data from, say, five (for larger schools) to 10 (for smaller schools) years may be accumulated to arrive at school mean residuals, with a mean being based on the most recent five or 10 years. As an example of how to start the process, Model I could be fitted to the data set comprised of the 1985-86, 1986-87, and 1987-88 freshman data. The derived equation could then be used to predict CCGPA for students from high schools that sent at least 10 students to UGA over the three years. Thus, the values of X_{12} through X_{16} would be determined, and Model J could be fitted to the combined data set and used for prediction with the 1989-1990 freshmen. Then to derive a rule for the 1989-90 freshmen, the same process could be repeated using classes of 1986-87, 1987-88, and 1988-89. (A three year period was used merely for illustration.)

Using Model J for Admission Decisions

As seen from the evaluation results, the mechanical use of Model J leads to, in general, unacceptable predictions. Although the bias in terms of mean residuals is acceptably low,

a residual standard deviation of about .53 is judged to be fairly large, indicating large variability and thus low predictive precision. More importantly, perhaps, is the consideration of the misclassification rates (see Table 11). The use of the prediction equation given our cut-off point of 2.00 resulted in a total error rate of about 18%. As we mentioned earlier, the effectiveness of the prediction equation in terms of the missclassification error rates is dependent on the cut-off value K, at which we artificially dichotimize the "success" and the "failure" groups. The relationship between the selection of K values and the resultant classification error rates is shown in Figure 1. The error rates were obtained

---

Insert Figure 1 about here

---

by applying Model J rule established on 1987-88 freshman data to the 1986-87 freshman data. We recommend that the effectiveness of the prediction equation be examined with reference to the values of K. The administrators can decide on the specific K value which yields the desired error rates according to the specific needs of their institution.

We also recommend that two alternative predicted CCGPA cut-off values be considered. One cut-off value, K_1 is the predicted CCGPA such that any prospective student with predicted CCGPA less than K_1 could be rejected. The second cut-

off value, K_2, is a value such that any prospective student with predicted CCGPA greater than K_2 could be accepted. For those prospective students with predicted CCGPA between K_1 and K_2, more information should be requested; for example, a writing sample, extracurricular activities, recommendations, etc. Based on our predictions with Model J on the 1987-88 data, using K_1=2.50 and K_2=3.00, the false-negative rate under such consideration is 8.9% and false-positive rate is 8.5%. These error rates are judged to be tolerably low. Of course, the selection of K_1 and K_2 values should be consistent with the needs of the institution. If acceptably low error rates can only be obtained with wide discrepancy between K_1 and K_2, it is an indication of the ineffectiveness of the prediction model, because most students would fall in the range between K_1 and K_2. The prediction equation would then have lost its value. For the 1987-88 freshman class, there were 1714 (about 50%) such "in-betweeners" when K_1=2.50 and K_2=3.00.

Discussion

The use of multiple regression in assessing college admissions prediction model effectiveness has an inevitable limitation due to the restriction of range on the predictor and criterion scores. It is understandable that students with low SAT or HSGPA scores would unlikely be admitted by the university admission office. The restriction of range effect

would naturally underestimate the true relationship between the predictors and the criterion measure. Therefore, one must not rely too heavily on a single R^2 value to judge the goodness of an prediction model. The effectiveness of the prediction model should also be assessed directly by the extent of accuracy that the model can be used in identifying "successful" or "failing" candidates.

We recommend the misclassification error rate as a preferred criterion for evaluating the effectiveness of a prediction model over others for the following three reasons: (1) it is free from statistical distribution assumptions; (2) it makes most intuitive sense since our concern is not to predict the exact value of CCGPA; rather, we want to predict whether an individual will be successful in college or not; and (3) it allows administrators to implement specific values of mininum satisfactory CCGPA (K) to meet the specific needs of the institution. By examining the relationship between indices of this type and the K values, one may have an idea of how effective the prediction equation functions. Based on our results, the best prediction model (Model J) functions unsatisfactorily in terms of the predictive precisions (absolute residual mean, mean square error, and the misclassification rate). Willingham (1990) has casted doubts over finding any additional measures that would add information of practical significance in predicting CCGPA, once HSGPA and SAT scores have been considered. All the evidence suggests that

the use of a prediction equation should be handled with great caution, and perhaps should not be used as the sole criterion for making admission decisions. If a prediction equation is to be constructed, we suggest that some more basic considerations be given in discovering, developing, and validating a potentially acceptable college admissions prediction model:

1. A variety of models be examined;

2. Definitions of indicator variables (such as our X12-X16) be regularly updated;

3. Regression weights used in the prediction model be regulary updated; and

4. Predictor and criterion measures be regularly examined for outliers and influential cases.

Another way of assessing predictive accuracy of a set of admission predictors may be accomplished via predictive discriminant analysis (PDA) -- see Huberty and Wisenbaker (in press) for examples. In PDA, the criterion variable is categorical. For the college admissions prediction problem, categories are imposed on the criterion variable, college academic performance -- in our situation, this was measured by CCGPA. In fact, analogous to the selection of a K value in the

regression prediction equation, predictive accuracy may be assessed for multiple definitions of criterion variable-imposed categories. As an example of multiple PDAs that may be conducted, we formed two groups of freshmen using 13 CCGPA cut-offs: 1.8, 1.9, ..., 2.9, 3.0; thus, 13 two-group PDAs were conducted. With each two-group PDA, false-positive, false-negative, and total misclassification rates can be obtained. The relationship of these error rates and K values can then be presented as in Figure 1. We established the classification functions in the context of PDA using the 1987-88 data, and validated the effectiveness of the classification functions (there are two classification functions in our case) with the 1986-87 data, the results are similar to those reported by classifications based on regression models. The adminintrators may wish to set "threshold" values for making admission decisions -- analogous to setting K_1 and K_2 using the regression equation for students defined as "in-betweeners." The posterior probabilities for assigning each student to the "success" and "failure" groups based on the observed predictor scores can be obtained in PDA. For example, those students who were assigned to groups with posterior probabilities less than a threshold of .60 belong to an "other" group; more information may be required of the students in this category. Depending on the K value selected, the posterior probability threshold could also be adjusted accordingly.

References

Arvey, R. D., & Faley, R. H. (1988). Fairness in selecting employees. Reading, MA: Addison-Wesley.

Baird, L. L. (1983). Predicting predictability: The influence of students and institutional characteristics on the prediction of grades. College Board Report No. 83-5.

Breland, H. M., (1979). Population validity and college entrance measures. Research Monograph No. 8. New York, NY: College Board.

Educational Testing Service. (1980). Test Use and Validity, Princeton, NJ: ETS.

Ford, S. F., & Campos, S. (1977). Summary of validity data from the admissions testing program validity study service. Princeton, NJ: College Entrance Examination Board.

Gatewood, R. D., & Feild, H. S. (1987). Human resource selection. Chicago: The Dryden Press.

Huberty, C. J, & Wisenbaker, J. M (in press). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2). Greenwich, CT: JAI Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Joreskog, C. G., & Sorbom, D. (1986). LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares estimates. Mooresville, IN: Scientific Software, Inc.

Ott, M. D., (1988). An analysis of predictors of early academic dismissal. Research in Higher Education. 28, 34-48.

Pedhazur, E. J. (1982). Multiple regression in behaviral research. New York: Holt, Rinehart & Winston.

Russell, J. S., Persing, D. L., Dunn, J. A., & Rankin, R. J. (1990). Focusing on the selection process: A field study of the measures that infuence admission decisions. Educational and Psychological Measurement, 50, 901-913.

SAS Institute Inc. (1985). SAS Users Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc.

Slack, W. V., & Porter, D. (1980). The scholastic aptitude test: A critical appraisal. Harvard Educational Review, 50, 382-391.

Terenzini, P. T. (1982). Designing attrition studies. In E. T. Pascarella (Ed.), New Directions for Institutional Research: Studying Student Attrition, No. 36. (pp. 55-71). San Francisco: Jossey-Bass.

Weitzman, R. A. (1982). The prediction of college achievement by the scholastic aptitude test and the high school record. Journal of Educational Measurement, 19, 179-191.

Willingham, W. W. (1990). Introduction: Interpreting predictive validity. In W. W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), Predicting college grades: An analysis of institutional trends over two decad. Princeton, NJ: Educational Testing Service.

Table 1.

Prediction Models Considered

Model	Criterion	Predictors
A	Z*Z	X1, X2, X3
B	Z	X1, X2, X3
C	Z	X1, X2, X4
D	Z	X1, X2, X4, X5
E	Z	X1, X2, X4, X6, X7, X8, X9
F	Z	X1, X2, X4, X1*X4, X2*X4
G	Z	X1*X1, X2*X2, X3*X3, X1, X2, X3
H	Z	X1, X2, X3, X10, X11
I	Z	X1, X2, X4, X10, X11
J	Z	X1, X2, X4, X10, X11, X12 through X16
K	-Z	X1, X2, X4, X10, X11, X12 through X16, X4*X12, X4*X13, X4*X14, X4*X15

Variable Definitions

Z	=	CCGPA
X1	=	SATV
X2	=	SATM
X3	=	Adj. HSGPA
X4	=	Unadjusted HSGPA
X5	=	High school GPA adjustment factor (.00, .20, .30, .40, ≥.50)
X6-X9	=	Indicator variables for 5 categories of X5
X10	=	Gender (male, female)
X11	=	Race (black, nonblack)
X12-X16	=	Indicator variables for identifying high school groups based on mean residuals (see text)

Table 2.

Descriptive Information for Four Measures Used in Model C

(1987-88 Data)

					Intercorrelations		
Measure	MEAN	STD DEV	MINIMUM	MAXIMUM	SATM	HSGPA	CCGPA
SATV	488.03	79.00	290.00	780.00	.43	.23	.33
SATM	539.08	78.49	290.00	780.00		.26	.31
HSGPA	3.02	.51	1.45	4.00			.52
CCGPA	2.53	.67	0.00	4.00			

* N = 3378

Table 3.

Model C Internal Residual Analyses, 1987-88 Data

	N	Raw Residual Mean	Absolute Residual Mean	Residual Std
Total Group	3378	.00	.43	.55
Male	1490	-.04	.45	.57
Female	1888	.03	.42	.53
Black*	162	-.15	.50	.60
Non-Black	3214	.01	.43	.55
HSGPA Adj. Factor				
.00	1363	-.06	.43	.55
.20	842	-.02	.45	.57
.30	437	.02	.42	.53
.40	352	.11	.42	.52
.50	317	.11	.43	.54
.60	67	.20	.42	.50
$\hat{Y} < 2.3$	1087	.03	.45	.57
$2.3 \leq \hat{Y} \leq 2.8$	1478	-.03	.43	.55
$2.8 < \hat{Y}$	813	.02	.42	.53

* Two students did not indicate race affiliation.

Table 4.

Internal Error Rates (in Percents) of Model C (for K=2.00)

(1987-88 Data)

	N	False Positive	False Negative	Total Error Rate
Total Sample	3378	15.8	54.2	18.2
Male	1490	19.3	47.0	21.5
Female	1888	13.2	64.0	15.5
Black*	162	27.5	66.7	29.6
Non-Black	3214	15.3	53.6	17.6
HSGPA Adj. Factor				
.00	1363	17.6	100.0	17.6
.20	842	17.1	39.1	17.7
.30	437	14.5	46.0	17.2
.40	352	9.8	56.1	18.5
.50	317	13.3	57.4	21.8
.60	67	7.7	73.3	22.4
$\hat{Y} < 2.3$	1087	30.5	54.2	35.0
$2.3 \leq \hat{Y} \leq 2.8$	1478	13.9	0.0	13.9
$2.8 < \hat{Y}$	813	3.3	0.0	3.3

* Two students did not indicate race affiliation.

Table 5.

Model J Internal Residual Analyses, 1987-88 Data

	N	Raw Residual Mean	Absolute Residual Mean	Residual Std
Total Sample	3376	.00	.42	.53
Male	1489	.00	.43	.55
Female	1887	.00	.41	.52
Black *	162	.00	.48	.60
Non-Black	3214	.00	.41	.53
HSGPA Adj. Factor				
.00	1362	-.01	.42	.54
.20	841	.01	.43	.56
.30	437	-.01	.41	.51
.40	352	.01	.39	.49
.50	317	.02	.40	.52
.60	67	.05	.37	.48
$\hat{Y} < 2.3$	1127	.02	.44	.57
$2.3 \leq \hat{Y} \leq 2.8$	1391	-.02	.42	.53
$2.8 < \hat{Y}$	858	.01	.38	.50

* Two students did not indicate race affiliation.

Table 6.

Internal Error Rates (in Percents) of Model J (for K=2.00)

(1987-88 Data)

	N	False Positive	False Negative	Total Error Rate
Total Sample	3376	15.0	52.8	18.1
Male	1489	17.2	48.8	21.6
Female	1887	13.4	64.0	15.4
Black*	162	26.2	65.6	34.0
Non-Black	3214	14.5	51.2	17.3
HSGPA Adj. Factor·				
.00	1362	16.1	53.1	17.8
.20	841	15.4	53.2	18.9
.30	437	14.5	51.2	18.1
.40	352	11.1	40.5	14.2
.50	317	14.8	60.4	22.4
.60	67	9.8	66.7	14.9
$\hat{Y} < 2.3$	1127	30.7	52.8	36.2
$2.3 \leq \hat{Y} \leq 2.8$	1391	13.1	0.0	13.1
$2.8 < \hat{Y}$	858	2.6	0.0	2.6

* Two students did not indicate race affiliation.

Table 7.

Descriptive Information for Four Measures. 1986-87 Data

					Intercorrelations		
Measure	MEAN	STD DEV	MINIMUM	MAXIMUM	SATV	HSGPA	CCGPA
SATM	530.99	79.68	290.00	790.00	.43	.26	.32
SATV	486.40	79.33	240.00	800.00		.22	.34
HSGPA	3.03	.51	1.65	4.00			.55
CCGPA	2.54	.66	0.00	4.00			

* N = 3177

Table 8.

Model C External Residual Analyses, 1986-87 Data

	N	Raw Residual Mean	Absolute Residual Mean	Residual Std
Total Sample	3177	.01	.41	.53
Male	1353	-.03	.42	.55
Female	1824	.03	.40	.51
Black	181	-.07	.38	.51
Non-Black	2996	.01	.41	.53
HSGPA Adj* Factor				
.00	1316	-.04	.42	.55
.20	712	-.03	.39	.51
.30	523	.03	.40	.52
.40	377	.12	.39	.49
.50	131	.10	.41	.54
.60	97	.18	.40	.49
$\hat{Y} < 2.3$	990	.00	.41	.54
$2.3 \leq \hat{Y} \leq 2.8$	1396	.00	.41	.53
$2.8 < \hat{Y}$	791	.02	.40	.51

* Twenty one students were from high schools with HSGPA adjustment factor of .70.

Table 9.

External Error Rates (in Percents) of Model C (for K=2.00)

(1986-87 Data)

	N	False Positive	False Negative	Total Error Rate
Total Sample	3177	14.2	56.5	16.6
Male	1353	16.8	54.6	19.5
Female	1824	12.3	58.8	14.4
Black	181	20.2	61.5	23.2
Non-Black-	2996	13.8	56.1	16.2
HSGPA Adj.* Factor				
.00	1316	15.6	60.0	15.8
.20	712	14.4	61.1	16.7
.30	523	13.3	51.4	15.9
.40	377	10.3	56.3	16.2
.50	131	17.7	58.6	26.7
.60	97	7.7	47.4	15.5
$\hat{Y} < 2.3$	990	28.3	56.5	33.3
$2.3 \leq \hat{Y} \leq 2.8$	1396	12.5	0.0	12.5
$2.8 < \hat{Y}$	791	2.8	0.0	2.8

* Twenty one students were from high schools with HSGPA adjustment factor of .70.

Table 10.

External Residual Analyses of Model J

(1986-87 Data)

	N	Raw Residual Mean	Absolute Residual Mean	Residual Std
Total Sample	3177	.00	.41	.53
Male	1353	.02	.43	.56
Female	1824	-.01	.40	.52
Black	181	.07	.40	.51
Non-Black	2996	-.00	.41	.53
HSGPA Adj.* Factor				
.00	1316	.00	.43	.55
.20	712	-.03	.41	.52
.30	523	.03	.41	.52
.40	377	.01	.38	.49
.50	133	-.04	.44	.57
.60	97	.06	.38	.51
$\hat{Y} < 2.3$	1044	.04	.42	.54
$2.3 \leq \hat{Y} \leq 2.8$	1300	.00	.42	.54
$2.8 < \hat{Y}$	833	-.04	.39	.51

* Twenty one students were from high schools with HSGPA adjustment factor of .70.

Table 11.

External Error Rates (in Percents) of Model J (for K=2.00)

(1986-87 Data)

	N	False Positive	False Negative	Total Error Rate
Total Sample	3177	13.9	62.2	17.9
Male	1353	16.0	62.6	22.3
Female	1824	12.4	61.3	14.6
Black	181	18.8	67.6	28.7
Non-Black-	2996	13.6	61.3	17.2
HSGPA Adj.* Factor				
.00	1316	15.0	65.2	16.8
.20	712	13.0	61.6	18.0
.30	523	13.0	66.2	20.0
.40	377	12.0	58.8	16.2
.50	131	18.5	56.5	25.2
.60	97	11.1	56.3	18.6
$\hat{Y} < 2.3$	1044	17.9	62.2	36.5
$2.3 \leq \hat{Y} \leq 2.8$	1300	12.2	0.0	12.2
$2.8 < \hat{Y}$	833	3.4	0.0	3.4

* Twenty one students were from high schools with HSGPA adjustment factor of .70.

Figure 1. Error Rates as a Function of K Values

Applying Model J Rule to 1986-87 Data